Alqalam Journal of Medical and Applied Sciences. 2025;8(3):1406-1411
https://doi.org/10.54361/ajmas.258322

Original article

Assessment of Agreement Among Evaluators in Scoring of Preclinical Full Metal Crown Preparation Using Global and Analytical Evaluation Methods

Nourelhouda Misurati*, Noora Berhaim

Department of Prosthodontics, Faculty of Dentistry and Oral Surgery, University of Tripoli, Libya.
Corresponding Email. *n.misurati@uot.edu.ly*

Abstract
It is essential to develop a reliable system for assessing preclinical tooth preparation in dental faculties. Many researchers have documented efforts to create dependable assessment systems for both laboratories and clinics. However, despite varying degrees of success, only a limited number of studies have employed an analytical approach to pinpoint the elements of the evaluation system that, with further refinement, could enhance reliability. This study aimed to compare the agreement among evaluators in two assessment methods: glance and grade (global) and checklist- and criteria-based assessment (analytical), for the evaluation of full metal crown preparation on the maxillary and mandibular first molars performed by 4th-year dental students. Sixty-nine fourth-year dental students and two faculty members from the fixed prosthodontics department, University of Tripoli, participated in this study. Each student prepared an ivorine tooth for the full metal crown on a maxillary or mandibular first molar on an upper or lower jaw dentoform. The two evaluators evaluated the work separately, and each evaluator first evaluated the preparation by glance and grade (global), and then by checklist and criteria-based assessment (analytical), the procedure was given a score on a 1 to 20 scale. The study revealed that, among the two evaluators, the level of intra-evaluator variability was not statistically significant. On the other hand, there were statistically significant differences in the level of inter-evaluator variability among evaluators evaluating full metal crown preparations with similar patterns of disagreement in both evaluation methods. It was concluded that among the two evaluators, there was inter-evaluator variability in pre-clinical evaluation using both global and analytical methods. On the other hand, the intra-evaluator variability was non-significant.
Keywords. Full Metal Crown Preparation, Dental Student's Assessment, Inter-Evaluator Variability, Intra-Evaluator Variability.

Introduction

Complete cast metal crowns are generally recommended for the restoration of severely damaged posterior teeth. The dentist's ability to adequately prepare the teeth is fundamental to the success of these restorations [1]. The British Society for Restorative Dentistry (1999) has outlined the principles of tooth preparation as follows: preservation of tooth structure, control of the path of insertion, optimum retention, and resistance form, appropriate occlusal clearance and articulation, and the removal of adequate tooth tissue to allow the manufacture of restorations with appropriate esthetic results without the over-contouring of the finished restoration [2].

The tooth preparation must have specific geometric properties to ensure the necessary retention and resistance to the vertical and lateral forces acting on the restoration. The presence of two opposing vertical surfaces is the most important element of retention. The axial walls of the preparation should be slightly tapered to allow the cementation of the artificial crown. The retention increases as the axial walls become more parallel. However, it is impossible to get parallel surfaces without creating undercuts [3].

Shillingburg et al, recommend a convergence angle (CA) between two opposing prepared axial surfaces of about (4°–6°) as ideal and a range of (4°–14°) as acceptable [1]. Goodacre et al suggest an angle of convergence between 10 and 20°. Also, Rosenstiel et al, suggest that [4,5]. Additionally, the occluso-cervical length is another fundamental factor affecting both retention and resistance. The longer the preparation, the greater the retention. Teeth with larger diameters require a greater length to prevent dislodgement [4].

Traditionally, tooth preparations have been assessed either visually by subjectively awarding a single grade or objectively by further dividing a single grade into several small grades. The researchers have named these two traditional methods as global grading (glance and grade method) and analytical grading (using rubrics). Recently, more sophisticated techniques have been presented, such as digitally grading tooth preparations using a variety of scanners and software programs introduced by various computer-aided design/computer-aided manufacturing (CAD/CAM). These techniques have been employed in practice in several institutions [6].

Habib S, suggests that the use of analytical rubrics in pre-clinical practice assessment provides appropriate feedback on the student's strengths and weaknesses in each parameter of their practical work, and also a consistency in evaluations was observed whilst using rubrics [7]. Evaluators' inconsistency may lead students to feel that assessment methods are somewhat arbitrary. Therefore, the objective of the current study was to compare the evaluator's agreements in two assessment methods: glance and grade (global),

Received*: 14-05-2025 -* ***Accepted****: 08-07-2025 -* ***Published****: 15-07-2025*

and checklist and criteria-based assessment (analytical) for the assessment of full metal crown preparation in grades awarded to the fourth-year students. The study is potentially useful for the development of a reliable system for the evaluation of the pre-clinical tooth preparation course in dental faculties.

Methods

The study was conducted at the Department of Prosthodontics, Faculty of Dentistry and Oral Surgery, University of Tripoli, Libya.

The study involved sixty-nine fourth-year dental students, and two staff members from the department of prosthodontics participated in this study. The students prepared an ivory tooth for the full metal crown on the maxillary or mandibular first molar on an upper or lower jaw dentoform. The students received clear instructions on how to prepare the full-metal crown and were informed about the evaluation criteria for the preparation.

Both evaluators are faculty members with Master's degrees in Fixed Prosthodontics and have been practicing and teaching Fixed Prosthodontics for over eight years.

The student's work was collected and given a numerical code. The two evaluators evaluated the work separately, with each evaluator first evaluating the preparation by glance and grade (global), and then by checklist and criteria-based assessment (analytical), the procedure was given a score on a 1 to 20 scale.

The preparations were blindly evaluated by each of the two evaluators (A, B) without magnification. For the initial assessment, each evaluator graded the preparation with the twenty-point scale using the eyeballing (glance and grade) method. After completion of the initial evaluation, a specific criterion for full metal crown preparation and an extended checklist, together with magnification, were used to verify the tooth preparation and the condition of the adjacent teeth. An analytic (criteria-based) assessment of various tooth preparation parameters was used for the second assessment by the 2 participating evaluators blindly.

The evaluators independently assessed the students' preparation in their own free time. There is no time limit specified for the assessment. However, the evaluators were requested to do the assessments alone and not in groups, and each student's score was recorded on a separate sheet by the two evaluators.

The criteria used in the study were based on a 20-point scale for evaluating 10 parameters of full-metal crown preparation. The score of each major parameter was further subdivided into a maximum score of 2, 1, and a minimum score of zero. Each of the criteria (parameters) was scored individually, and the rightmost column was filled with the specific score for each parameter. The sum of all scores for each parameter was treated as a student's total score (Table 1).

Table 1. Criteria for evaluation of full metal crown preparation.

Criteria	Excellent (2)	Satisfactory (1)	Unacceptable (zero)
Occlusal reduction	Normal anatomic contours are followed, 1 mm on non-functional cusps and 1.5 mm on functional cusps.	Underprepared, less than the recommended, but can be corrected.	Over-reduction, more than the recommended, but cannot be corrected.
Buccal reduction	0.5 mm of uniform reduction is achieved. (Minimum 0.5mm, maximum 0.7mm)	Under-prepared, less than the recommended, but can be corrected.	Over-reduction, more than the recommended but cannot be corrected.
Functional cusp bevel	Apparent bevel, placed at an angle of approximately 45 degrees to the long axis of the prepared tooth.	Under-prepared functional cusp bevel	Absence of functional cusp bevel.
Lingual reduction	0.5 mm of uniform reduction is achieved. (Minimum 0.5mm, maximum 0.7mm)	Under-prepared, less than the recommended, but can be corrected.	Over-reduction, more than the recommended but cannot be corrected.
Proximal reduction	0.5 mm of uniform reduction is achieved. (Minimum 0.5mm, maximum 0.7mm)	Under-prepared, less than the recommended, but can be corrected.	Over-reduction, more than the recommended but cannot be corrected.
Axial taper reduction	Parallel axial wall with minimal taper	(Slight over-tapering) presence of taper, but near parallel or over-tapered on the mesial or distal wall	(Severe over-tapering) on more than two axial walls
Presence of undercuts	No undercuts.	Presence of a slight undercut on one or two axial walls.	Presence of undercuts on more than two axial walls.
Rounded line angles	All line angles are rounded and smooth	Slightly rounded angles, or the presence of sharp angles.	Overall rough preparation with apparent sharp angles

Received: *14-05-2025* - ***Accepted***: *08-07-2025* - ***Published***: *15-07-2025*

Location of the chamfer finish line	Finish line at the gingival margin, clear, smooth, and continuous	Finish line slightly sub- or supra-gingival, clear, smooth, and continuous	Clear supra-gingival or deeply sub-gingival, unclear, irregular, not continuous.
Preservation of the adjacent teeth	Adjacent teeth are unaffected	Adjacent teeth are minimally touched.	Adjacent teeth are abraded and flattened.

Statistical Package for Social Sciences (SPSS) software version 26 was used for the analysis of the obtained data. The significance level was set at P≤0.05. A descriptive statistic for quantitative data was done (mean, median, standard deviations, and range) for both Evaluation Methods conducted by the two evaluators. A Cohen's Kappa test was used to measure the level of agreement between the two evaluators evaluating full metal crown preparation on maxillary and mandibular first molars performed by fourth-year dental students using a criteria-based checklist. Kappa value interpretation is as follows: values ≤ 0 indicate no agreement, from 0.01- 0.20 slight, 0.21- 0.40 fair, 0.41- 0.60 moderate, 0.61- 0.80 substantial, and 0.81- 1.00 as almost perfect agreement (Landis & Koch, 1977). To assess the level of agreement (Intra-rater reliability) among each evaluator, the interclass correlation (ICC) was used.
The paired t-test, and the nonparametric Wilcoxon Signed-Ranks Test (non-parametric alternative to paired t-test), were used for the comparison among quantitative data, by calculating the differences and testing whether they differ from zero between the two evaluator's scores using "Glace and Grade assessment" and the "checklist criteria-based assessment", P-value < .05 was considered at level of significance.

Results

Descriptive statistics on student scores awarded by the two evaluators are presented in Table 2. Cohen's Kappa Statistic was run to determine if there was evaluation agreement between the two evaluators evaluating full metal crown preparation on the maxillary and mandibular first molar performed by 69 fourth-year dental students. Cohen's Kappa values result for (Buccal reduction, Functional cusp bevel, Lingual reduction, Axial taper reduction, Rounded line angles, and Location of chamfer finish line) shows that the level of agreement (K) was fair (K= 0.23, 0.30, 0.37, 0.25, 0.22, 0.29) respectively, which is statistically significant with P-value ranged between (<.0001 - <.008). While the level of agreement between two evaluators in (Occlusal reduction k = 0.51, Proximal reduction k = 0.4, and Preservation of the adjacent teeth k = 0.52, Presence of undercuts k= 0.5) shows that the level of agreement (K) was moderate, which is considered statistically significant with P-value <.0001

Table 2. Descriptive statistics of student grades awarded by the two examiners in two evaluation methods (n = 69):

Evaluation Method	Examiner	Minimum	Maximum	Median	Mean ± Std.D
Criteria-based checklist assessment (analytical)	Examiner 1	5	19	13	12.93 ± 3.46
	Examiner 2	2	18	12	11.72 ± 3.33
Glance and grade "visual inspection" (global)	Examiner 1	5	18	14	13.91 ± 3.02
	Examiner 2	3	18	13	12.80 ± 3.54

Std.D. Standard Deviation.

The ICC was used to evaluate the intra-evaluator reliability agreement between the scores assigned by global glance and grade assessment and checklist criteria-based assessment of two evaluators, Table (3). Shows that intra-rater reliability was excellent. 0.93, with a 95% confidence interval of the difference (0.78 - 0.97) for examiner 1, and 0.92, with a 95% CI (0.77 - 0.96) for examiner 2.

Table 3. Intraclass correlation Agreement between the global "glance and grade" assessment and the checklist criteria-based assessment:

Comparison		95% CI		
	ICC [a]	Lower Bound	Upper Bound	Sig
Conventional glance and grade assessment and checklist criteria-based assessment of examiner 1	.93	.78	.97	<.0001
Conventional glance and grade assessment, and checklist criteria-based assessment of examiner 2	.92	.77	.96	<.0001

ICC, Intraclass Correlation, CI, Confidence Interval of the Difference Two-way mixed effects model where people effects are random and measures effects are fixed.[a.] Type intraclass correlation coefficients using an absolute agreement definition.

Received*: 14-05-2025 -* ***Accepted****: 08-07-2025 -* ***Published****: 15-07-2025*

a-Glance and grade assessment of Evaluator 1 /glance and grade assessment of Evaluator 2. The paired t-test demonstrated a statistically significant difference between glance and grade assessment of Evaluator 1 (13.91 ± 3.02) and Evaluator 2 (12.80 ± 3.53), t (68) = 4.034, P < .0001. Table (4). b-Checklist criteria-based assessment of Evaluator 1 / Checklist criteria-based assessment of Evaluator 2:
A Wilcoxon signed-rank test (non-parametric test) was run to determine if there is a difference between the scoring of the two evaluators using checklist criteria-based assessment (analytical rubrics), the results indicate a statistically significant difference between evaluator 1, with a median and standard deviation of (13 ± 3.46) and evaluator 2 (12 ± 3.33), (Z = -5.020, P < .001).

Table 4. Paired t-test assessing the difference between two examiners' ratings using global "glance and grade" assessment (n = 69):

Pair	Paired Differences						
				95% CI			
	Mean	Std.D	Std.EM	Lower	Upper	t	Sig. (2-tailed)
Global assessment of Evaluator 1- Evaluator 2	1.12	2.30	.28	.56	1.67	4.034	<.0001

Std.D. Standard Deviation, Std.EM, Standard Error Mean, CI, Confidence Interval of the Difference.

Discussion

In accordance with the standard prosthodontics literature and textbooks, proper and detailed preparation of the tooth prior to the fabrication of a full-coverage crown should be carried out to produce a clinically successful prosthesis [1,4,5]. The principles of tooth preparation, which include appropriate retention and resistance, structural durability, preservation of tooth structure and periodontium, and marginal integrity, should be followed during crown preparation [1]. Accurate and fair evaluation of student work is usually considered the most crucial stage of education. In preclinical dental education, students need to receive consistent and accurate feedback from the faculty in order to achieve a higher level of performance before moving on to the clinics [8].
In this research study, two assessment methods: glance and grade (global), and checklist and criteria-based assessment (analytical rubrics) were followed for the evaluation of full metal crown preparation done by fourth-year dental students in pre-clinical fixed prosthodontics. Among the two evaluators, the level of intra-evaluator variability was not statistically significant. On the other hand, there were statistically significant differences in the level of inter-evaluator variability among evaluators evaluating full metal crown preparations with similar patterns of disagreement in both evaluation methods (global vs analytical). In many teaching institutions and due to practical situations, the glance and grade method is still applied, especially with more experienced faculty staff. In our study, inter-evaluator variability was detected among the two evaluators with the global method, Evaluator 1 (13.91 ± 3.02) and Evaluator 2 (12.80 ± 3.53), t (68) = 4.034, P < .0001. Something that had previously been reported both by Jenkins et al. [9], using a global evaluation method, and AlHumaid et al. [10], using a rating scale that did not include descriptions of the levels of performance.
The application of a criteria-oriented grading system in dentistry dates back more than four decades. Dhuru et al. [11] 1978, emphasized the importance and recommended the use of criteria-oriented grading for pre-clinical dentistry. The usefulness of this criteria-based evaluation has been highlighted by many researchers in the following years [6,7,12,13]. The results of our study showed a difference between the grades of the two evaluators using analytical rubrics (inter-evaluator variability), the results indicate a statistically significant difference between evaluator 1, with a median and standard deviation of (13 ± 3.46) and evaluator 2 (12 ± 3.33), (Z = -5.020, P < .001). These findings are similar to a previous study conducted by Sharaf et al. [14], and Al Amri et al. [15], who indicated that the problem of inter-examiner reliability and variability still exists even with using analytical rubrics for evaluation.
Many studies suggest that the use of analytic rubrics could increase the consistency of grades amongst the evaluators [6,13,16,17]. However, our study did not agree with previous studies like Habib [6], which used an analytical system for evaluation using specific criteria and a checklist in an attempt to reduce variability among examiners. He reported that the technique was better than the glance and grade method in reducing variability among examiners. Also, Satheesh et al. [13], reported increased reliability of 90.2% with the use of analytic rubrics in their research study, and Escribano et al. [16], suggested that the use of an analytic rubric allowed different evaluators to reach higher levels of agreement than those obtained with a numeric rating scale in the evaluation.
Based on the results of this study, variations between the evaluators were observed within the various parameters, especially for (buccal reduction, functional cusp bevel, lingual reduction, axial taper reduction, rounded line angles, and location of chamfer finish line) which showed that the level of agreement (K) was fair (K= 0.23, 0.30, 0.37, 0.25, 0.22, 0.29) respectively, which is statistically significant with P-value ranged between (<.0001 - <.008). While the level of agreement between two evaluators for (occlusal reduction k =

Received*: 14-05-2025 -* ***Accepted****: 08-07-2025 -* ***Published****: 15-07-2025*

0.51, proximal reduction k = 0.4, and preservation of the adjacent teeth k = 0.52, the presence of undercuts k= 0.5) was moderate, which is considered statistically significant with P-value <.0001. The present study showed that intra-evaluator reliability was excellent. [0.93, with a 95% confidence interval of the difference (0.78 - 0.97) for evaluator 1, and 0.92, with a 95% CI (0.77 - 0.96) for evaluator 2]. This observation is in agreement with Sharaf et al. [14], who also found non-significant intra-examiner variations in most preparations, and in disagreement with Jenkins et al. [9], opposed their conclusion who found significant intra-examiner variability. The evaluator's consistency is crucial in the teaching and learning process as it can affect the confidence and performance of the students. Therefore, new evaluation techniques and methods of standardizing assessments need to be further studied to promote an efficient system of learning [14]. The use of computer evaluation is an approach that may increase agreement among evaluators. In this case, preparations are electronically scanned and compared to gold standard standardized samples using specially developed software that generates numerical values for agreement.[18] These interventions need to be tested in future studies for their impact on improving evaluators' agreement. Many studies suggest that using digital grading will preclude the variability and subjectivity that usually result from traditional visual inspection grading [19-23].

Conclusion

It was found that Inter-evaluator variability in preclinical tooth preparation evaluation exists in both global and analytical methods among the two examiners. In contrast, the intra-evaluator variability was found to be non-significant. The utilization of analytical rubrics for the assessment of various success criteria of full metal preparations by dental students is valuable and can help in finding the inaccuracies and deficiencies within each parameter of their work. Despite utilizing analytical rubrics, variations in the evaluator's grades still existed. Highlighting the need for further research and improvement of rubrics. Application of new evaluation methods and tools in Libyan dental schools is required, such as a digital grading system using software programs to evaluate the prepared teeth by the three-dimensional (3D) technique. Application of this approach as a part of dental education will preclude the subjectivity and variability that usually result from the traditional visual inspection grading, and reduce the workload of faculty teaching members.

Conflict of interest. Nil

References

1. Shillingburg HT, Hobo S, Whitsett LD, Jacobi R, Brackett SE. Fundamentals of fixed prosthodontics. 4th Edition, Chicago, Hanover Park, IL: Quintessence Publishing Company; 2012.
2. British Society for Restorative Dentistry. Guidelines for crown and bridge. *Eur J Prosthodont Restor Dent.* 1999; 7(1): 3-9.
3. Rosella D, Rosella G, Brauner E, Papi P, Piccoli L, & Pompa G. A tooth preparation technique in fixed prosthodontics for students and neophyte dentists. *Annali di stomatologia.* 2015; 63(4):104-119
4. Goodacre CJ, Compagni WV, Aquilino SA. Tooth preparations for complete crowns: An art form based on scientific principles. *J Prosthet Dent.* 2011; 85(4):363-376.
5. Rosenstiel, S.F., Land, M.F. and Walter, R.D. Contemporary fixed prosthodontics. 6th Edition, Philadelphia, Elsevier Health Sciences; 2023.
6. Habib, S.R. Rubric system for evaluation of crown preparation performed by dental students. *Eur J Dent Edu. 2018; 22(3)*: 506–513.
7. Habib, S.R. Assessment of inter-examiner variability and criteria for post and core preparation by analytical rubrics. *Pakistan Oral & Dental Journal.* 2020; 40(3):190-196.
8. Tahani, B., Rashno, A., Haghighi, H., Monirifard, R., Khomami, H. N., & Kafieh, R. Automatic Evaluation of Crown Preparation Using Image Processing Techniques: A Substitute to Faculty Scoring in Dental Education. *Journal of medical signals and sensors*, 2020;*10* (4), 239–248.
9. Jenkins SM, Dummer PMH, Gilmour ASM, Edmunds DH, Hicks R, Ash P. Evaluating undergraduate preclinical operative skill; use of a glance and grade marking system. J Dent. 1998;26(8):679–84.
10. AlHumaid J, Tantawi M, El, Al-Ansari AA, Al-Harbi FA. Agreement in Scoring Preclinical Dental Procedures: impact on grades and instructor-related determinants. *J Dent Educ.* 2016;80(5):553–62.
11. Dhuru VB, Rypel TS, Johnston WM. Criterion-oriented grading system for preclinical operative dentistry laboratory course. *J Dent Educ.* 1978; 42(9):528-531.
12. Allen D, Tanner K. Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sci Educ.* 2006; 5(3):197-203.
13. Satheesh KM, Brockmann LB, Liu Y, Gadbury-Amyot CC. Use of an analytical grading rubric for self-assessment: a pilot study for a periodontal oral competency examination in predoctoral dental education. *J Dent Educ.* 2015; 79(12):1429-1436.
14. Aly A. Sharaf, Amr M. AbdelAziz, Omar A.S. El Meligy. Intra- and inter-examiner variability in evaluating preclinical pediatric dentistry operative procedures. *J Dental Edu.* 2007; 71(4): 540-544.
15. Al Amri, M.D, Sherfudhin H.R, Habib, S.R. Effects of evaluator's fatigue and level of expertise on the global and analytical evaluation of preclinical tooth preparation. *J of Prostho* 2016; 27(7);1-8
16. Escribano N, Belliard V, Baracco B, Da Silva D, Ceballos L, M. Victoria Fuentes. Rubric vs. Numeric Rating Scale: Agreement among Evaluators on Endodontic Treatments Performed by Dental Students. *BMC Medical Education* 2023;23(97):1-9

Received: *14-05-2025* - ***Accepted***: *08-07-2025* - ***Published***: *15-07-2025*

17. Alammari MR, Alkhiary YM, Nawar AA. Intra-and inter-examiner variability in evaluating impression procedures at the undergraduate level. *J Life Sci.* 2013;5(1):5–10.
18. Renne WG, McGill ST, Mennito AS, et al. E4D compare software: an alternative to faculty grading in dental education. *J Dent Educ* 2013;77(2):168-75.
19. Kateeb, E.T., Kamal, M.S., Kadamani, A.M., Abu Hantash, R.O., & Abu Arqoub, M.M. Utilising an innovative digital software to grade pre-clinical crown preparation exercise. *Eur J Dent Educ.* 2017;21(4);220–227.
20. Miyazono, S., Shinozaki, Y., Sato, H., Isshi, K., & Yamashita, J. Use of Digital Technology to Improve Objective and Reliable Assessment in Dental Student Simulation Laboratories. *J dent educ.* 2019;83(10):1224-1232
21. Schepke, Ulf; van Wulfften Palthe, Mariëtte E; Meisberger, Eric W; Kerdijk, Wouter; Cune, Marco S; Blok, Berend. Digital Assessment of a Retentive Full Crown Preparation—An Evaluation of PrepCheck in an Undergraduate Pre-clinical Teaching Environment. *Eur J Dent Educ.*2020;24(10):417-424
22. Liu J, Liu Y, Wang J, Zuo X, Wang X, Zhang Y, et al. Dental measurements based on a three-dimensional digital technique: a comparative study on reliability and validity. *Arch Oral Biol.* 2021; 124:105059.
23. Azhari A, Ahmed W, Almaghrabi R, Almalki A, Merdad Y, Alrefai H, Bahrawi M, Kordi L. Comparing the reliability of inter- and intra-grader using digital scanning vs. traditional visual method for evaluating preclinical class II composite preparation. *The Saudi Dent J.* 2024;36(9):1253-1259.

Received*: 14-05-2025 -* ***Accepted****: 08-07-2025 -* ***Published****: 15-07-2025*